DRIVES
Development and Research on Innovative Vocational Education Skills

U3 MACHINE LEARNING ALGORITHMS

U3.E4 REINFORCEMENT LEARNING MODELS

Machine Learning Engineer

January 2021, Version 1

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

The student is able to

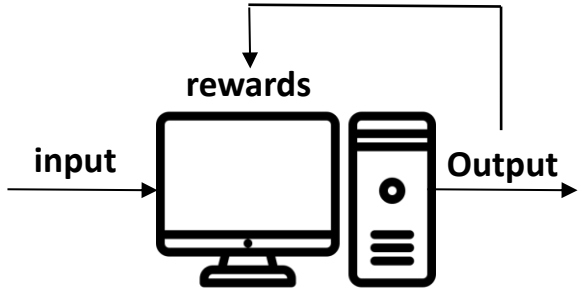
MLE.U3.E4.PC1	Define and understand reinforcement learning.
MLE.U3.E4.PC2	Know common terminologies used in the field of reinforcement learning.
MLE.U3.E4.PC3	Know some of the most commonly used algorithms in reinforcement learning.
MLE.U3.E4.PC4	Understand model-based reinforcement learning and model-free reinforcement learning as well as their differences.
MLE.U3.E4.PC5	Understand policy optimization methods, Q-learning methods and hybrid methods as well as list some algorithms for each category.
MLE.U3.E4.PC6	Understand the different approaches of model-based reinforcement learning: learn the model and learn given the model.
MLE.U3.E4.PC7	Know the domain application of reinforcement learning models.

MACHINE LEARNING ALGORITHMS

SUPERVISED LEARNING

UNSUPERVISED LEARNING

REINFORCEMENT LEARNING

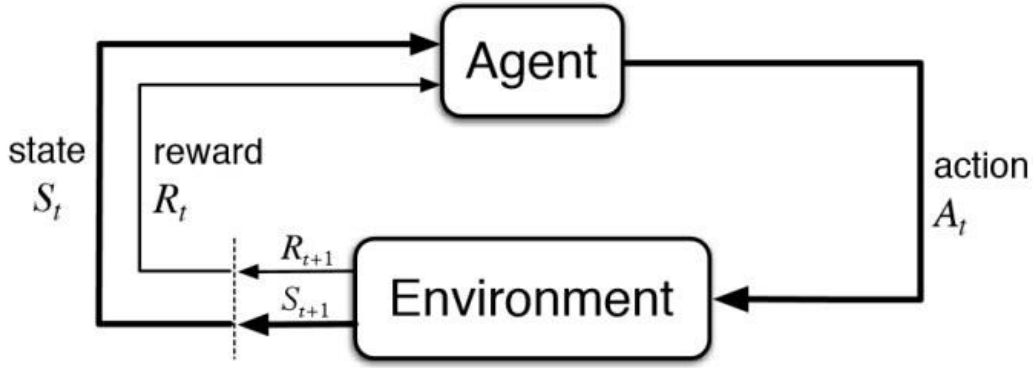


REINFORCEMENT LEARNING

Does not require labeled input/output pairs to be submitted

It is a paradigm that is concerned with how software agents should act in an environment in order to maximize the notion of cumulative reward.

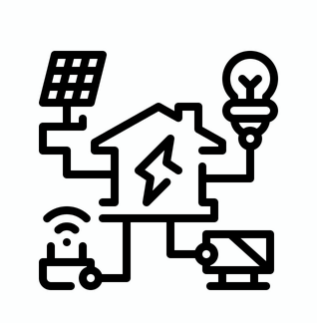
It is a type of dynamic programming that forms algorithms using a reward and punishment system



TERMINOLOGY

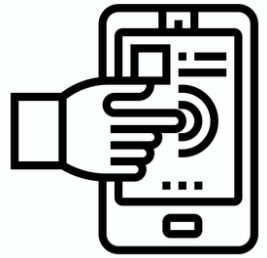
Agent

the learner and the decision maker.



Environment

Where the agent learns and decides what actions to perform.



Action

A set of actions which the agent can perform.

State

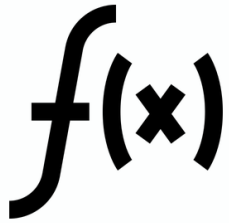
The state of the agent in the environment.

Reward

For each action selected by the agent the environment provides a reward. Usually a scalar value.

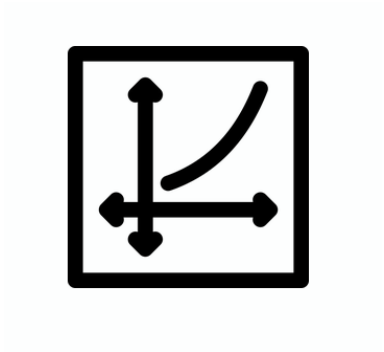
Policy

The decision-making function (control strategy) of the agent, which represents a mapping from situations to actions.



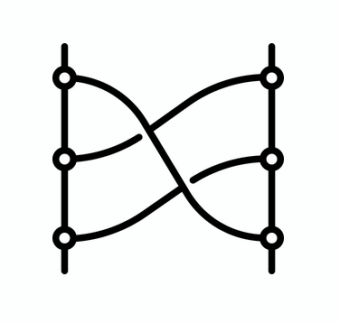
Value function

Mapping from states to real numbers, where the value of a state represents the long-term reward achieved starting from that state, and executing a particular policy.



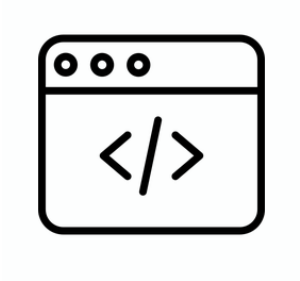
Function approximator

Refers to the problem of inducing a function from training examples. Standard approximators include decision trees, neural networks, and nearest-neighbor methods



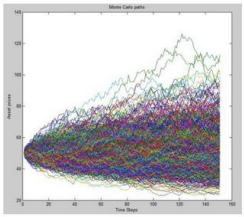
Markov decision process (MDP)

A probabilistic model of a sequential decision pro blem, where states can be accurately perceived, a nd the current state and action chosen to determine the probability distribution of future st ates. Essentially, the outcome of applying an actio n to a state depends only on the current action an d state (and not on previous actions or states).



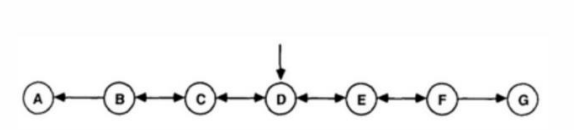
Dynamic programming (DP)

is a class of solution methods for solving sequential decision problems with a compositional cost structure. Richard Bellman was one of the principal founders of this approach.



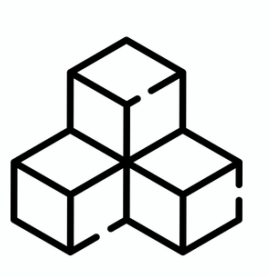
Monte Carlo methods

A class of methods for learning of value functions, which estimates the value of a state by running many trials starting at that state, then averages the total rewards received on those trials.



Temporal Difference (TD) algorithms

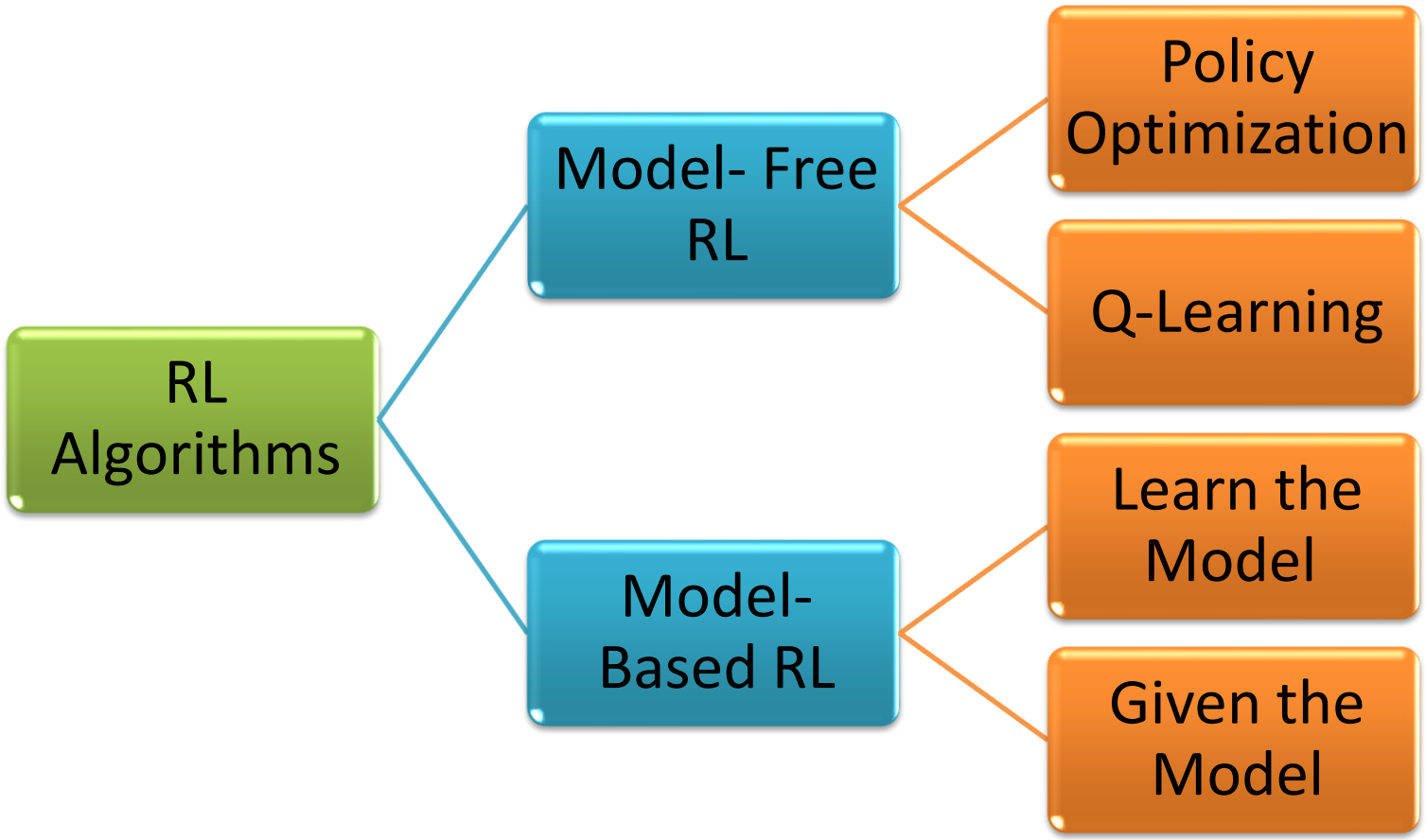
A class of learning methods, based on the idea of comparing temporally successive predictions. Possibly the single most fundamental idea in all of reinforcement learning.



Model

The agent's view of the environment, which maps state-action pairs to probability distributions over states. Note that not every reinforcement learning agent uses a model of its environment

TYPES OF REINFORCEMENT LEARNING ALGORITMS



MODEL-FREE

The agent relies on trial-and-error experience to establish the optimal policy.

It is statistically less efficient than model-based methods

The transition probability distribution and reward function are very often referred to as the "model" of the environment.

There are two main approaches to representing actors according to this type of learning:

- Policy Optimization
- Q-Learning

DRIVES
Development and Research on Innovative Vocational Education Skills

Policy Optimization Principals

- The agent directly learns the political function that maps the state of action.
- The policy is determined without using a value function.
- There are two types of policies: deterministic and stochastic.
- The **deterministic** policy maps declare for action without uncertainty.
- Stochastic policies produce a probability distribution over shares in each state. This process is called the Partially Observable Markov Decision Process (POMDP).

DRIVES
Development and Research on Innovative Vocational Education Skills

Q-learning Principals

- Learns the action-value function Q(s, a): how good it is to take action in a given state.
- For any Markov finite decision-making process (FMDP), Q-learning finds an optimal policy.
- "Q" designates the function that returns the reward used to provide the booster.
- There are at least 2 widely used variants: Deep Q-Learning and Double Q-Learning.

DRIVES
Development and Research on Innovative Vocational Education Skills

STEP BY STEP

01 Initialize Q Table

02 Choose an action a

03 Perform the action

04 Measure Reward

05 Update Q

06 Back to Step3

DRIVES
Development and Research on Innovative Vocational Education Skills

Policy optimization

- Policy gradient
- A2C/A3C: Asynchronous Advantage Actorcritic
- PPO: Proximal Policy Optimization
- TRPO: Trust Region Policy Optimization

Q-learning

- DQN: Deep Q Neural Network
- C51
- QR-DQN: Distributional Reinforcement Learning with Quartile Regression
- HER: Hindsight Experience Replay

MODEL BASED

Uses experience to develop an internal model of transitions and immediate results in the environment

The goal is to plan through a control function f(s,a) to choose the optimal actions

The appropriate actions are then chosen through research or planning in this global model

There are two main approaches to representing:

Learning the Model

Learn by given the model

MODEL BASED:LEARN THE MODEL

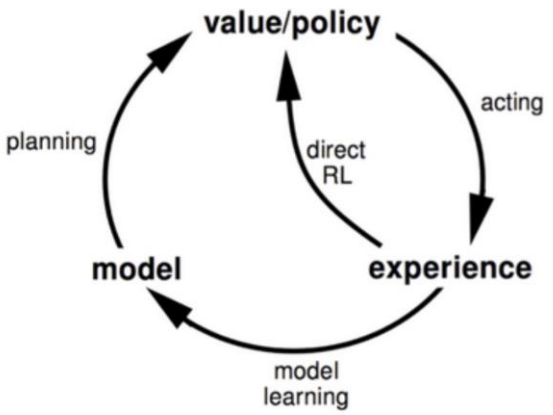
To learn the model, a basic policy is executed while the trajectory is followed. The model is stored using the sampled data.

STEP BY STEP:

01 Run Base Policy to collect data

02 Learn Dynamic model f(s,a) to minimize the summatory

03 Plan through f(s,a) to choose actions



DRIVES
Development and Research on Innovative Vocational Education Skills

Learn the models

- World models
- I2A: Imagination-Augmented Agents
- MBMD: Model Based Priors for Model Free Reinforcement Learning
- MBVE: Model Based Value Expansion

Given the Models

- AlphaZero

HYBRID ALGORITHMS

DDPG: Deep Deterministic Policy Gradients

SAC: Soft Actor Critic

TD3: Twin Delayed Deep Deterministic Policy Gradients

REINFORCEMENT LEARNING: ADVANTAGES

DRIVES
Development and Research on Innovative Vocational Education Skills

- Can be used to solve very complex problems that cannot be solved by conventional techniques;
- This technique is preferred to achieve long-term results, which are very difficult to achieve.
- It is very similar to the learning of human beings. Therefore, it is close to achieving perfection.
- Can correct the errors that occurred during the training process.
- Once an error is corrected by the model, the chances of occurring the same error are very less.
- Robots can implement reinforcement learning algorithms to learn how to walk.
- Reinforcement learning models can outperform humans in many tasks
- Reinforcement learning is intended to achieve the ideal behavior of a model within a specific context, to maximize its performance.
- It can be useful when the only way to collect information about the environment is to interact with it.

REINFORCEMENT LEARNING: DISADVANTAGES

- Reinforcement learning as a framework is wrong in many different ways, but it is precisely this quality that makes it useful.
- Too much reinforcement learning can lead to an overload of states, which can diminish the results.
- It is not preferable to use for solving simple problems.
- Needs a lot of data and a lot of computation.
- Assumes the world is Markovian, which it is not.

REINFORCEMENT LEARNING APPLICATIONS

Resources management in computer clusters

Designing algorithms to allocate limited resources to different tasks is challenging and requires human-generated heuristics.

Self-Driving Cars

Various papers have proposed Deep Reinforcement Learning for autonomous driving. In self-driving cars, there are various aspects to consider, such as speed limits at various places, drivable zones, avoiding collisions

Industry Automation

In industry reinforcement, learning-based **robots** are used to perform various tasks. Apart from the fact that these robots are more efficient than human beings, they can also perform tasks that would be dangerous for people.

REINFORCEMENT LEARNING APPLICATIONS

DRIVES
Development and Research on Innovative Vocational Education Skills

Trading And Finance

Supervised time series models can be used for predicting future sales as well as predicting stock prices. However, these models don't determine the action to take at a particular stock price.

Healthcare

In healthcare, patients can **receive treatment** from policies learned from RL systems. RL can find optimal policies using previous experiences without the need for previous information on the mathematical model of biological systems.

News Recommendation

User preferences can change frequently, therefore **recommending news** to users based on reviews and likes could become obsolete quickly. With reinforcement learning, the RL system can track the reader's return behaviors.

SUMMARY

- Reinforcement Learning does not require labeled input/output pairs to be submitted;
- RL has its own terminology, it is very important for understanding the model
- There are 2 types of RL algorithms: Model Free and Model Based
- Model free has 2 main approaches: Policy optimization and Q learning
- Model Based also has 2 main approaches: Learning the model and learning by a given model
- The 2 types have their advantages and disadvantages as well as their use cases

REFERENCE TO AUTHORS

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0003-2326-2153

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0002-2988-196X

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-4121-6169

REFERENCE TO AUTHORS

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0001-6457-0756

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **“Machine Learning Engineer”:**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES

Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.